

**CLAIMS**

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

- 1        1. A computer implemented method for categorizing incoming electronic  
2        communications using a supervised machine learning component, where the  
3        method factors an organization's business domain into the technology domain  
4        to enable an acceptable automated response and routing system, said method  
5        comprising the steps of:
  - 6                (a) analyzing the business domain;
  - 7                (b) determining an approach to machine learning in the form of a  
8                program or an algorithm that will be used to induce a categorizer using  
9                supervised learning, the categorizer being generated from training data  
10               comprising a set of examples of the communications of a kind to be classified;
  - 11               (c) collecting existing data of representative examples of electronic  
12               communications and inventories of personnel skills, business processes,  
13               workflows, and business missions;
  - 14               (d) analyzing the collected data, thereby gaining an appreciation of the  
15               complexity, vagueness, and uniqueness to be expected in the communications  
16               to be categorized, as well as the relative numbers of various kinds of  
17               communications, and also thereby determining a technical structure of the  
18               communications that is likely to be relevant to categorization, as well as,  
19               factoring the inventories of personnel skills, business processes, workflows,  
20               and business missions collected, providing a basis for obtaining a complete  
21               understanding of what must be done with each electronic communication, and  
22               by whom.
  - 23               (e) defining a categorization scheme;

24 (f) labeling examples of electronic communications with categories  
 25 from the categorization scheme for use both as training data to be used in the  
 26 supervised learning step and as test data;

27 (g) converting the labeled data into a form suitable for subsequent  
 28 processing, both for purposes of machine learning and technical validation;

29 (h) performing machine based supervised learning technology to  
 30 induce a categorizer for the categorization scheme; and

31 (i) validating the categorization scheme with respect to technical  
 32 performance and business requirements.

1 2. A method as recited in claim 1, further comprising the step of  
 2 implementing the categorization scheme by putting the categorization system  
 3 into production.

1 3. A method as recited in claim 1, further comprising the steps of:  
 2 reviewing the categorization scheme to consider its adequacy in light  
 3 of recent distribution of communications; and  
 4 modifying the categorization scheme, as required, to accommodate  
 5 new business goals, or to keep in step with changes in the supervised learning  
 6 technology, wherein if it is determined to change the categorization scheme,  
 7 steps (f) through (i) are repeated.

1 4. A method as recited in claim 1, wherein the step of analyzing the business  
 2 domain further comprises steps:  
 3 analyzing anticipated content of relevant electronic communications;  
 4 analyzing business missions and goals;  
 5 evaluating skills of involved personnel;  
 6 analyzing the organization's workflow;

7 analyzing use of stored responses including determining whether  
 8 answers have been developed for frequently occurring questions; and  
 9 producing business requirements for use in the validation step using  
 10 insight gained by the analysis of the business domain.

1 5. A method as recited in claim 4, wherein the step of analyzing business  
 2 missions and goals further comprises:  
 3 reviewing a model of the business domain and determining success  
 4 criteria and measurements used to determine when the business is successful;  
 5 establishing turnaround times for the electronic communications to  
 6 support mission and goals of the business; and  
 7 determining a volume of electronic communications received daily and  
 8 determining a number of received communications that must be answered to  
 9 meet the goals of the business.

1 6. A method as recited in claim 4, wherein the step of analyzing the  
 2 organization's workflow, further comprises:  
 3 determining a workflow through the organization and routing  
 4 performed on a category by category basis;  
 5 determining if subject matter experts (SME) have been established for  
 6 categories of information; and  
 7 determining whether an automated or manual system for routing  
 8 electronic communications is being used.

1 7. A method as recited in claims 1, wherein the step of defining a  
 2 categorization scheme further comprises steps:  
 3 combining lists of categories in the group of categories related to  
 4 business mission groups, related to routing communications to specific

5 individuals, communications for which an automated response is feasible and  
6 desirable, and those related to existing stored responses or stored templates  
7 for responses;  
8 determining technically feasible categorization from the assembled  
9 categorization scheme;  
10 correlating knowledge of the technical structure of the communications  
11 with knowledge of what kinds of features can actually be identified by the  
12 machine learning component; and  
13 eliminating or combining categories with few examples, if necessary.

1 8. A method for categorizing incoming electronic communications using a  
2 supervised machine learning component, where the method factors an  
3 organization's business domain into the technology domain to enable an  
4 acceptable automated response and routing system, said method comprising  
5 the steps of:  
6 selecting a machine learning component for the technology domain;  
7 preparing a set of training data comprising representations of  
8 previously categorized electronic communications, wherein the data in an  
9 electronic communication is textual and each electronic communication has  
10 features, where a feature is related to textual data;  
11 analyzing the organization's business domain with respect to desired  
12 routing and handling of contemplated message categories of electronic  
13 communications, the analysis resulting in identification of tasks to be  
14 performed and actions to be taken in response to a received electronic  
15 communication of a contemplated message category, the analysis also  
16 resulting in identification of features relevant to categorization of electronic  
17 communications;  
18 determining skill levels of personnel corresponding to required tasks

19 and actions identified in the step of analyzing the organization's business  
20 domain;  
21 extracting a new representation of each electronic communication in  
22 the training set depending on a frequency of occurrence in the electronic  
23 communication of features identified as relevant to the business domain;  
24 inducing a pattern characterization when an electronic communication  
25 belongs to a category, wherein the patterns may be presented as rules or  
26 another format corresponding the selected machine learning component; and  
27 developing an initial categorization scheme based on areas of the  
28 business domain receiving a greater quantity of electronic communications or  
29 electronic communications of a relatively higher priority.

1 9. A method as recited in claim 8, wherein an electronic communication  
2 comprises more than one part and each part of the electronic communication  
3 has corresponding features related to a category and may be categorized based  
4 on each part in the inducing step.

1 10. A computer implemented method for routing electronic communications  
2 where a categorization scheme is determined by analyzing an organization's  
3 business domain with respect to desired routing and handling of contemplated  
4 message categories of electronic communications, said method comprising the  
5 steps of:  
6 representing a newly arrived electronic communication in a  
7 predetermined same format as a training set of data used to train a supervised  
8 machine learning component of an electronic communications categorizer;  
9 inputting the representation of the newly arrived electronic  
10 communication to the categorizer;  
11 identifying features in the newly arrived electronic communication;

00280621AA

12 matching identified features with features selected during a “feature  
13 selection” phase of training the supervised machine learning component,  
14 wherein feature selection is performed using business domain information of  
15 an organization as a factor;  
16 determining at least one category for the newly arrived electronic  
17 communication based on the matched features; and  
18 routing the newly arrived electronic communication to at least one  
19 person in the organization, where the communication is routed based on tasks  
20 and actions identified as necessary for the organization's business domain  
21 which correspond to the at least one determined category.